1974

Signal detection theory :

Joseph C. Weber
Lehigh University

SIGNAL DETECTION THEORY
A Comparison of Alternative Models

by

Joseph C. Weber

A Thesis

Presented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Psychology

Lehigh University

1974

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

5/2/74
(date)

Edwin J. Kay
Professor in Charge

Arthur J. Brody
Chairman of Department

Acknowledgements

I would like to dedicate this thesis to my wife Bobbi. Without her many hours of data collecting, typing and moral support, this work would not have been possible.

I would also like to thank Dr. Edwin J. Kay, whose ideas, suggestions and criticisms are in evidence on every page of this thesis.

I am also indebted to Dr. Martin L. Richter for his invaluable aid in helping to analyze the results and interpret the data.

Table of Contents

List of Tables

List of Figures

Abstract

Five basic models, derived from Signal Detection Theory (SDT), were formulated to describe the decision process of a subject in detecting signal in a noisy background. Their differential predictions of the subject's performance were compared in terms of a modified four alternative forced-choice (4AFC) procedure in which one or two or three observations within a trial may contain signal. The threshold model was compared with four integration models differing along two dimensions: (a) amount of information used in forming the decision (Total or Partial Integration), and (b) the method employed by the subject in forming the decision (Majority or Absolute Decision Rule). Two trained observers were used to generate the data and a 2AFC procedure was used before and after each test session to determine the subject's level of detection.

Multiple χ^2 comparisons indicated that the threshold model was clearly superior in predicting the subject's performance in the modified 4AFC procedure. Attempts to use the threshold model in predicting the subject's percentage of correct responses (p(c)) on the modified 4AFC procedure from the results on the 2AFC procedure for each test session were not successful. The predicted p(c) values were consistent overestimates of the actual p(c)

data. A predicted relationship of the threshold model, i.e., p(c) increases as the number of observations of signal (n) within a trial increases, was upheld. In general, the threshold model predictions were reliable estimates of the actual data that was generated.

Introduction

Signal Detection Theory (SDT) is a psychophysical theory used to explain how an "ideal" subject detects signals in a noisy backround. In the general SDT procedure, a subject observes either noise alone (N) or signal-plus-noise (S). The information he receives from a stimulus is summarized internally by a real number. From repeated observations of the two stimuli, S and N, two random variables are produced by the subject, one arising from observations of signal (X_S) and one from observations of noise (X_N). Both are normally distributed with variance σ^2. The noise distribution has mean $E(X_N) = \mu_N$ and the signal distribution has mean $E(X_S) = \mu_S$. The effect of adding signal to noise is to shift the signal distribution to the right of the noise distribution by distance $d = \mu_S - \mu_N$. These assumptions of SDT describe the information processing or receptive stage of the hypothetical subject.

Further assumptions are necessary to describe the decision process of the subject. The subject is considered to be an "ideal observer" or a "Baysian calculator", for in order to make a decision, the observer must know the parameters of the distributions for S and N and he must know the a priori probabilities of S and N. With this information, the ideal observer will set a criterion value (c) and compare with it the number he internalizes from an individual observation (X) in such a way that:

if $X > C$, he will respond S;

if $X < C$, he will respond N.

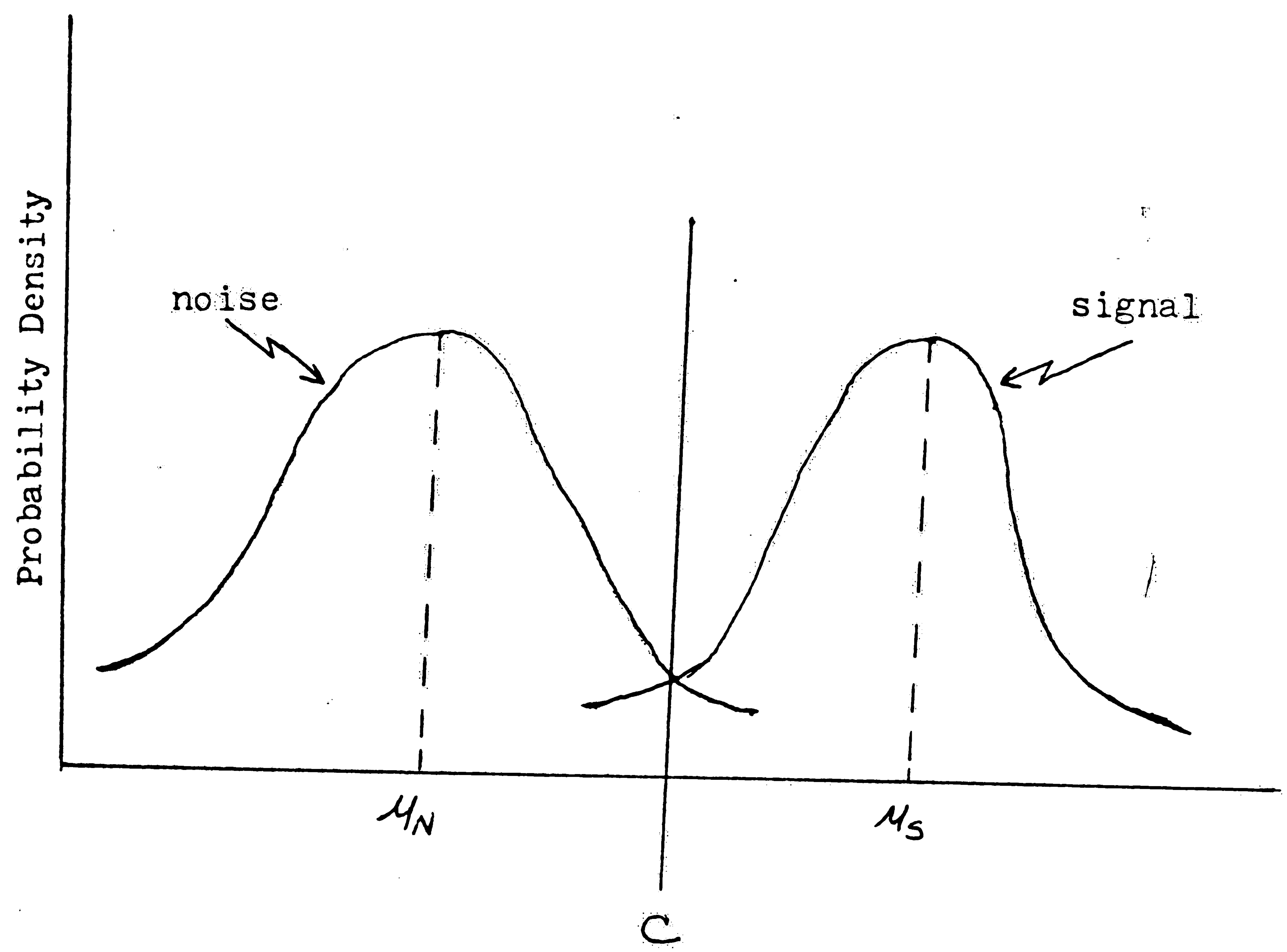
The value of this criterion value (C) is a function of the *a priori* probabilities, the parameters of the distributions and the subject bias--all of which the ideal observer knows. Therefore, if there is an equal likelihood of S and N and if there is no advantage of or bias for correctly identifying S rather than N, the subject will place the criterion value such that it will maximize detection of S and N. Figure 1 illustrates where the criterion value would be set within the two distributions of S and N.

Two basic tasks are generally used in SDT research. In the Yes-No task (YN), the subject is presented with a single stimulus and must decide whether it was S or N. The Forced-Choice procedure (FC) is different in that the subject is presented with a number of stimulus alternatives and must decide which one of them is signal. For example, in a two-alternative forced-choice procedure (2AFC), two stimuli are presented: one signal and one noise. The subject must decide which one was S. There appears, then, to be only a quantitative difference between these two procedures since a NAFC task can be viewed as N Y-N tasks with the single restraint of only one possible Yes re-

Figure 1

Distributions of S and N

with criterion value(C)


Probability Density
noise
signal
μ_N
μ_S
C

sponse. It is possible, however, to ask the subject to make a decision based on multiple observations of the test stimulus, rather than just one. This is more than just a quantitative difference in procedure, since it introduces the question of how the subject combines the information he receives from all the observations he has made to form a single decision. Two possible models will be presented to explain how the final decision is made.

The Integration model is sometimes called the "detection-theory" model since it is most compatible with SDT. The model assumes that the ideal observer has perfect memory for the real numbers used in summarizing the information he receives from an observation. There is no loss of information with increasing observations. This means that the information the subject receives from the first observation in a trial is preserved over all successive observations. The real numbers used to summarize the information received from each observation are added together to form the accumulation of evidence that is used in making the final decision. To show that the assumption of perfect memory can, in fact, hold up experimentally, a detectability index $d' = (\mu_S - \mu_N) / \sigma$ is computed. d' measures the distance between the signal and noise distributions in standard deviation units. The farther apart they are, the easier they are to discriminate. Swets (1959) found that observers operated with the same efficiency (yielded essentially the same d') regardless of the number of alternatives. Since the calculation of d' assumes perfect memory, those results imply that the observer is capable of storing and selecting among eight measures obtained in a 8AFC task just as well as when fewer alternatives are offered. The integration model further assumes that observations are independent (i.e., the information received from one observation is not a factor of any previous observation) and that each observation is normally distributed with variance σ^2. Therefore, a special value of d' can be measured for each observation. It has been found that, in general, d' increases by a factor of the square root of the number of observations: $d'_n = \sqrt{n}d'$ (Swets, et. al., 1959).

The Threshold Model, on the other hand, does not assume perfect memory. It is assumed that only the decision after each observation is retained and information from previous observations is lost. Therefore, the overall decision is solely a function of the individual decisions that are made from each observation. Independence of observations is assumed, thereby inferring an increase in detectability with increased observations. Each independent observation presents another independent detection opportunity. As the number of observations (n) increases, the probability of detection increases: $p_n = 1 - (1 - p)^n$, where p is the probability of detection on a single observation. A positive response (S) is made for the overall decision if any one of the multiple observations is positive (i.e., if any one of the sensory measures exceeds a threshold). The sensory thresholds fluctuate over time and the threshold levels are assumed independent from one observation to another.

The primary goal of this study is to compare the predictions of the threshold model along with various forms of the integration model. In order to do this, a modification of the 4AFC procedure is used which allows for multiple observations within a single trial. As indicated, in the 4AFC procedure, the subject is presented with four stimuli and must decide which one of them is signal. The modified procedure, however, allows for one, two or three of the stimuli to include signal. The subject, then, is presented with two different sources of information, channel A and channel B, one of which always contains signal, and the other, noise. A total of four observations of these two channels, in various orders are offered in each trial: 3 from A and 1 from B; 3 from B and 1 from A; or 2 from A and 2 from B. The subject must decide which channel, A or B, contains the signal. This procedure, therefore, generates the possibility of any one of 14 distinct combinations of S and N on any given trial. Table 1 illus-

Table 1

Trial Sequences of 4 Observations

of S and N

OBSERVATION

TRIAL SEQUENCE	1	2	3	4
1	S	S	S	N
2	S	S	N	S
3	S	N	S	S
4	N	S	S	S
5	S	S	N	N
6	S	N	N	S
7	N	N	S	S
8	S	N	S	N
9	N	S	N	S
10	N	S	S	N
11	N	N	N	S
12	N	N	S	N
13	N	S	N	N
14	S	N	N	N

trates how the 14 possible trial sequences are formed over four observations.

The integration model, as indicated above, assumes that the information from observations of a channel is summed to form a decision. However, the integration model may be further characterized by the amount of information used and the method in which this usable information is employed to make the decision. Total integration assumes that the information from all observations of a channel is used in the decision process. Partial integration assumes that information from only the last sequence of observations of a channel is incorporated into the decision process. The intervening observations are assumed to interfere with the integration of information, and the information is discarded if not used for a decision. A decision can be made by comparing the information obtained from channel A (a_1, a_2, a_3,...a_n) to that obtained from channel B (b_1, b_2, b_3,...b_m). A majority decision rule is defined as a rule by which the subject decides that the signal is statistically more likely to be in channel A if:

$$\sum_{i=1}^{n} a_i/n \geqslant \sum_{j=1}^{m} b_j/m.$$

An absolute decision rule is one in which the subject compares both sets of information to a criterion value (c), and decides that the signal is more likely to be in channel A if: $\sum_{i=1}^{n} a_i \,/n \geq c$ and $\sum_{j=1}^{m} b_j \,/\, m \leq c$.

The information from both sources, S and N, is assumed to be integrable with equal ease. Of course, if only one source of information is being used to make the decision, an absolute decision rule must be employed by the subject. Also, if:

$$\sum_{i=1}^{n} a_i / n \geq c \text{ and} \sum_{j=1}^{m} b_j / m \geq c$$

or if:

$$\sum_{i=1}^{n} a_i / n \leq c \text{ and} \sum_{j=1}^{m} b_j / m \leq c$$

the subject simply guesses.

Four distinct integration models are offered as possible models used in a subject's decision process. Their predictions are discussed along with those of the threshold model, in terms of the multiple observation 4AFC procedure outlined above.

I. Total Majority

The total majority model assumes that information from all observations of a channel is used and a majority decision rule is employed by the subject. Information from both sources, S and N, is integrated and compared to make the decision. The subject, therefore, using this model, will maximize his chances of being correct when the most information from both sources is available for comparison. This should occur when equal numbers of observations of S and N are presented on a trial. To illustrate why a subject should be more successful on trials containing 2S and 2N rather than three observations of one channel and one of the other, an examination of the variance resulting from these two types of trials is most descriptive. Given 4 observations in a single trial, X, Y, W and Z, the information received on a trial with 3N + 1S or 3S + 1N will be:

$$Q = \frac{X + Y + Z}{3} - W$$

$$\sigma_Q^2 = 1/9(3)\sigma^2 + \sigma^2$$

$$= 4/3\sigma^2$$

since:

$$\sigma_X^2 = \sigma_Y^2 = \sigma_Z^2 = \sigma_W^2 = \sigma^2$$

Given that the trial consists of 2S and 2N, the information received becomes:

$$R = \frac{X + Y}{2} - \frac{W + Z}{2}$$

$$\sigma_R^2 = 1/4(2)\sigma^2 + 1/4(2)\sigma^2$$

$$= \sigma^2$$

Since $d' = (\mu_S - \mu_N)/\sigma$ is a function of the distance between the distribution means and the variance, it can be shown, Swets, et. al., (1959), that detectability decreases as the variance increases. Therefore, all trial sequences consisting of 2S + 2N (σ^2) should yield better detectability than any other trial sequence consisting of 3S + 1N or 3N + 1S, where the variance is larger ($4/3\sigma^2$). The model prediction, then, in terms of trial sequences listed in Table 1, can be stated:

[5,6,7,8,9,10] > [1,2,3,4,11,12,13,14]

where the trial sequences listed within a set of brackets ([]) are predicted to yield equivalent detection and the common mathematical symbol ($>$) can be read in this case to mean "will yield better detectability than."

II. Partial Majority

This model assumes partial integration of information along with a majority decision rule. Partial integration assumes that only the information from the last set of observations of a channel is used in the decision process. This allows for the possibility of only 3 or even 2 observations being used to form a decision. In continuing our examination of variance to include these possibilities, we discover that given 3 observations, 2N + 1S or 2S + 1N,

$$T = \frac{X + Y}{2} - Z$$

$$\sigma_T^2 = 2(1/4)\sigma^2 + \sigma^2$$

$$= 3/2\sigma^2$$

Given 2 observations, 1S + 1N,

$$V = X - Y$$

$$\sigma_V^2 = \sigma^2 + \sigma^2$$

$$= 2\sigma^2$$

Combining these results with those shown above, we find that the trial sequences can be rank ordered by degrees of variance in the following order:

2S + 2N	σ^2
3S + 1N or 3N + 1S	$4/3\sigma^2$
2S + 1N or 2N + 1S	$3/2\sigma^2$
1S + 1N	$2\sigma^2$

Since the majority decision rule is employed in this model, given the same total number of observations, equal numbers of observations per channel will be optimal for a subject's performance. However, the integration of information will be dependent upon where those observations are placed within a trial. Only the last sequence of observations of each channel will be integrated. It is clear, then, that trial sequences 5 and 7 (Table 1) will maximize a subject's performance, since all the information is integrated and there are equal numbers of observations from each channel. Trial sequences 1, 4, 11 and 14 also allow for all observations to be integrated, but there are unequal numbers of observations per channel. The remaining trial sequences allow for either 3 or 2 observations to be integrated and the model thus predicts inferior performance by the subject when presented with these trial sequences based on their higher variance. The model prediction, therefore, becomes:

$$[5,7] > [1,4,11,14] > [3,6,10,13] > [2,8,9,12]$$

III. Partial--Last Channel

This model assumes a partial integration of information in which the subject takes the information from the last observation on a trial and makes a decision based on that single piece of information alone. Effectively, this becomes a YN task where the subject uses an absolute decision rule. If the information he receives is

$$(a \text{ or } b) > c,$$

he says signal, and if

$$(a \text{ or } b) < c$$

he says noise. All trial sequences yield equal predictions in this case since the information from the first three observations in each trial is discarded:

$$[1 = 2 = 3 = 4 = \ldots. = 14]$$

IV. Partial--Last Channel Sequence

This model assumes partial integration in which the subject makes a decision based only on the last channel of information observed. An absolute decision rule is again employed and the model prediction becomes a factor of the number of observations in the last channel presented. Trial sequences 4 and 14 have three observations in the last channel presented; trial sequences 3, 5, 7 and 13 have two observations and the remaining trial sequences contain only one. Therefore, the model prediction becomes:

$$[4,14] > [3,5,7,13] > [1,2,6,8,9,10,11,12]$$

V. Threshold

The threshold model differs from all forms of the integration model in that only the individual decisions after each observation (detect or non-detect) are used to make the final decision. In the threshold model, the subject "chooses" a channel if any observation of that channel leads to detection of signal. If no signal is detected, he guesses. Each observation from the S source adds another detection opportunity and therefore increases the subject's rate of performance. Clearly, the trial sequences with the most observations of signal will maximize the subject's probability of making a correct response. The model prediction, then, becomes a factor of the number of Ss in a trial:

$$[1,2,3,4] > [5,6,7,8,9,10] > [11,12,13,14]$$

Method

Subjects. Two trained observers, the experimenter and his wife, served as subjects. Training consisted of approximately 25 one-hour sessions of practice using a 2AFC procedure. Training continued until a criterion of at least 5 consecutive sessions of non-significant variability (as measured by a χ^2 test) in detection level was reached.

Apparatus. The experimental room was soundproofed and equipped with a comfortable chair opposite a wall on which were mounted two 24 volt lights, approximately 1 inch in diameter, one green and one red. Through the use of BRS Foringer digital logic equipment, the experimenter could turn on either the red or the green light for .7 seconds. Along with the presentation of a light, white noise was simultaneously presented over earphones with a 13 kHz tone present or absent. Robinson and Trahiotis (1972) compared the procedures using simultaneous onset and continuous noise conditions and found that the subject's performance improved as the delay between noise and signal onset increased. To avoid such differences as are found with continuous noise, a simultaneous noise-signal onset procedure was used. The subject also had the use of three response buttons mounted on the arm of the chair: one to start the trial, and two response keys, red and green.

The experiment consisted of five distinct parts:

Part I

Procedure. The subject was presented with 100 2AFC trials before each test session (pre-test) and immediately following each test session (post-test). These trials were used to determine the subject's detection level before and after each test session. The subject was presented with two channels of information, green followed by red, and his task was to select which channel contained signal. Before each test session, the subject was given a cue tone which gave him the opportunity to hear the signal before the test sequence began. No trial-by-trial feedback was employed. Gundy (1961) found that subjects who were permitted to hear the signal before the test sequence maintained a stable level of performance throughout the session in an SDT task while those who were not afforded this opportunity performed at chance level initially and gradually improved. Throughout the 100 trials, the channel containing signal was randomly assigned to red or green. A target detection level of 75% was considered optimal and a comparison of the pre-test and post-test scores for each session was used to give an indication of whether or not this detection level varied throughout the course of the session.

Results. Table 2 contains the pre-test and post-test scores of subjects B and J for each of the 7 test sessions of the experiment. A χ^2_1 was obtained for each set of scores. None of the χ^2_1's were significant, indicating that for each pair of scores, the pre-test scores did not differ significantly from the post-test scores.

Note: In this and in all other parts of the experiment, the data presented are pooled over 100 or 150 trials. Since the models presented, describing the psychological process in detecting a tone, are formulated in terms of events which occur on each presentation of a stimulus, this may appear to raise the question of the subject's response reliability to a particular stimulus. That is, can massing the data over trials tell us how the subject is responding to an individual stimulus? Madigan (1971) tested response reliability in a YN and 2AFC task and found that, in both cases, the mean responses obtained are reliable measures of detectability of individual stimuli.

Part II

Procedure. The modified 4AFC procedure, described above, was used to provide a test of the five models presented for comparison. Table 3 provides a summary of the differential predictions made by each of the models in terms of the trial sequences presented in Table 1. To

Table 2

Pre-test and Post-test scores
for subject B and J with χ^2_1 for sessions 1-7

	Subject					
	B			J		
Session	Pre-test	post-test	χ^2_1	Pre-test	post-test	χ^2_1
1	65	68	1.034	75	69	.892
2	83	83	0	59	54	.51
3	78	76	.106	78	82	.5
4	78	81	.276	82	75	1.452
5	74	78	.372	74	74	0
6	81	77	.394	62	62	0
7	65	63	.088	60	64	.34

N = 100

Table 3

Predictions of the models
presented for comparison

Model	Model Predictions
I. Total Majority	[5,6,7,8,9,10] > [1,2,3,4,11,12,13,14]
II. Partial Majority	[5,7] > [1,4,11,14] > [3,6,10,13] > [2,8,9,12]
III. Partial--Last Channel	[1=2=3=....=14]
IV. Partial--Last Channel Sequence	[4,14] > [3,5,7,13] > [1,2,6,8,9,10,11,12]
V. Threshold	[1,2,3,4] > [5,6,7,8,9,10] > [11,12,13,14]

compare the predictions of each model, 7 test sessions, each containing 4 different trial sequences, were run. The trial sequences were chosen in order to maximize the differential predictions of the models and provide a clear test between the different model predictions. Table 4 provides an outline of the trial sequences that were used in each test session along with the prediction each model makes about those particular sequences. For each test session, each sequence is presented 150 times, a total of 600 trials per test session. The order of trial sequences within a test session was completely random, as was the channel, red or green, containing the signal.

Results. Table 5 contains the scores of both subjects for each test session. A χ^2 analysis was performed on these data to test each model prediction. Table 6 shows the results of the χ^2 tests. In many cases, only one χ^2 test was needed to test a model prediction for a test session. In test session 1, for example, Models I, II and III predicted that trial sequences 1, 4, 11 and 14 would yield the same results. To test this prediction, a χ^2_3 was done to determine if the four data points were significantly different. Model IV, however, makes three predictions about the trial sequences of session 1: it predicts that sequence 4 will yield the same results as sequence 14; sequence 11 will yield the same results as sequence 1; and sequences 4 and 14 will yield better results than sequences

Table 4

Model Predictions of Results

for Each Test Session

		Model		
Session	Stimuli	I	II	III
1	1,4,11,14	Same	Same	Same
2	5,7,8,9	Same	[5,7] > [8,9]	Same
3	3,13,2,12	Same	[3,13] > [2,12]	Same
4	1,11,5,7	[5,7] > [1,11]	[5,7] > [1,11]	Same
5	6,10,5,7	Same	[5,7] > [6,10]	Same
6	4,14,6,10	Same	[4,14] > [6,10]	Same
7	6,10,3,13	Same	Same	Same

		Model	
Session	Stimuli	IV	V
1	1,4,11,14	[4,14] > [11,1]	[1,4] > [11,14]
2	5,7,8,9	[5,7] > [8,9]	Same
3	3,13,2,12	[3,13] > [2,12]	[2,3] > [12,13]
4	1,11,5,7	[5,7] > [1,11]	[1] > [5,7] > [11]
5	6,10,5,7	[5,7] > [6,10]	Same
6	4,14,6,10	[4,14] > [6,10]	[4] > [6,10] > [14]
7	6,10,3,13	[3,13] > [6,10]	[3] > [6,10] > [13]

Table 5

Number Correct for Subjects B and J

for All Test Sessions in Part II

Session	Trial Sequence	Subject B	Subject J
1	(1) SSSN	95	134
	(4) NSSS	112	128
	(11) NNNS	76	111
	(14) SNNN	55	106
2	(5) SSNN	93	95
	(7) NNSS	83	95
	(8) SNSN	94	94
	(9) NSNS	86	96
3	(3) SNSS	111	120
	(13) NSNN	108	115
	(2) SSNS	115	122
	(12) NNSN	99	102
4	(1) SSSN	103	128
	(11) NNNS	78	111
	(5) SSNN	83	126
	(7) NNSS	85	131
5	(6) SNNS	91	108
	(10) NSSN	86	119
	(5) SSNN	89	107
	(7) NNSS	93	118
6	(4) NSSS	117	132
	(14) SNNN	89	99
	(6) SNNS	98	121
	(10) NSSN	103	119
7	(6) SNNS	103	115
	(10) NSSN	102	126
	(3) SNSS	120	135
	(13) NSNN	82	102

N = 150

Table 6

χ^2 Tests of Model Predictions

on All Test Sessions

Stimuli	Model I	Model II
1,4,11,14	Same B J 49.03** 22.23**	Same B J 49.03** 22.23**
5,7,8,9	Same B J 2.38 .06	[5,7] > [8,9] B J [5,7] 1.37 0 [8,9] .89 .06 [5,7][8,9] .11 0
3,13,2,12	Same B J 4.61 9.00*	[3,13] > [2,12] B J [3,13] .15 .49 [2,12] 4.17* 7.05** [3,13][2,12] .21 1.12
1,11,5,7	[5,7] > [1,11] B J [5,7] .05 .68 [1,11] 8.71** 5.95* [5,7][1,11] 1.16 3.77	[5,7] > [1,11] B J [5,7] .05 .68 [1,11] 8.71** 5.95* [5,7][1,11] 1.16 3.77
6,10,5,7	Same B J .94 4.38	[5,7] > [6,10] B J [5,7] .22 2.15 [6,10] .34 2.19 [5,7][6,10] .17 .28
4,14,6,10	[6,10] > [4,14] B J [6,10] .38 .08 [4,14] 12.14** 20.49** [6,10][4,14] .19 .80	[4,14] > [6,10] B J [4,14] 12.14** 20.49** [6,10] .38 .08 [4,14][6,10] .19 .80
6,10,3,13	[6,10] > [3,13] B J [6,10] .01 2.55 [3,13] 21.88** 21.88** [6,10][3,13] .07 .17	Same B J 22.21** 25.06**

Stimuli	Model III	Model IV
1,4,11,14	Same B 49.03** J 22.23**	[4,14] > [11,1] B J [4,14] 43.88** 9.40** [11,1] 4.91* 11.78** [4,14][11,1] .11 1.25
5,7,8,9	Same B 2.38 J .06	[5,7] > [8,9] B J [5,7] 1.37 0 [8,9] .89 .06 [5,7][8,9] .11 0
3,13,2,12	Same B 4.61 J 9.00*	[3,13] > [2,12] B J [3,13] .15 .49 [2,12] 4.17* 7.05** [3,13][2,12] .21 1.12
1,11,5,7	Same B 9.78* J 11.07*	[5,7] > [1,11] B J [5,7] .05 .68 [1,11] 8.71** 5.95* [5,7][1,11] 1.16 3.77
6,10,5,7	Same B .94 J 4.38	[5,7] > [6,10] B J [5,7] .22 2.15 [6,10] .34 2.19 [5,7][6,10] .17 .28
4,14,6,10	Same B 12.59** J 22.52**	[4,14] > [6,10] B J [4,14] 12.14** 20.49** [6,10] .38 .08 [4,14][6,10] .19 .80
6,10,3,13	Same B 22.21** J 25.06**	[3,13] > [6,10] B J [3,13] 21.88** 21.88** [6,10] .01 2.55 [3,13][6,10] .07 .17

Stimuli	Model V		
1,4,11,14	[1,4] > [11,14]		
		B	J
	[1,4]	4.50*	1.81
	[11,14]	5.96*	.42
	[1,4][11,14]	39.13**	20.96**
5,7,8,9	Same		
		B	J
		2.38	.06
3,13,2,12	[2,3] > [12,13]		
		B	J
	[2,3]	.29	.37
	[12,13]	1.26	2.81
	[2,3][12,13]	2.99	5.79*
1,11,5,7	[1] > [5,7] > [11]		
		B	J
	[5,7]	.05	.68
	[1][5,7]	4.64*	.01
	[11][5,7]	.48	6.34*
	[1][5,7][11]	9.38**	9.37**
6,10,5,7	Same		
		B	J
		.94	4.38
4,14,6,10	[4] > [6,10] > [14]		
		B	J
	[6,10]	.38	.08
	[4][6,10]	4.23	3.57
	[14][6,10]	2.06	7.46**
	[4][6,10][14]	12.34**	21.64**
6,10,3,13	[3] > [6,10] > [13]		
		B	J
	[6,10]	.01	2.55
	[3][6,10]	5.33*	5.57*
	[13][6,10]	6.13*	5.61*
	[3][6,10][13]	22.05**	22.39**

11 and 1. χ^2 were performed to determine the accuracy of each prediction. In a like manner, each model prediction was tested for each set of trial sequences in the 7 test sessions. 58 χ^2 tests were performed overall. The χ^2 tests were not all independent in that identical model predictions were tested with the same χ^2 test. Table 7 contains the numbers of χ^2 tests used to test each model along with the number of tests which yielded accurate predictions. Table 8 shows the results of the χ^2 tests over each test session. Checkmarks (✓) are used to indicate which model(s) are supported for each test session. The last column in Table 8 indicates the subjective evaluation of the experimenter as to how much support was given to Model V over each test session.

Part III

Procedure. A possible deficiency in the procedure of Part I is that the order of channel predentations is not random, i.e., green was always presented first, followed by red. Thus the subject did not have to attend to the lights to identify which channel of information was being presented. Máloney and Welch (1972) found that the presence of an accessory tone decreased the percentage of light detections when the tone was loud and continuous. Tones of short duration and lower intensity were found to increase visual detectability. In order to determine if

Table 7

Number of χ^2 Tests Yielding

Predicted Results for Each Model

Model	#of χ^2 Tests	# of Predicted Results B	J
I	13	6	5
II	17	7	7
III	7	3	2
IV	21	8	8
V	20	14	18

Table 8

Models Supported by χ^2 Tests

for All Test Sessions in Part II

Session	Preferred Model I	II	III	IV	V	Evaluation of Support for Model V
1					✓	unequivocal
2	✓		✓		✓	reasonable
3					✓	qualified
4					✓	unequivocal
5	✓		✓		✓	reasonable
6					✓	unequivocal
7					✓	unequivocal

the presence of the channel indicator lights had any such effects on the detectability of a tone, Part III of the experiment was run, comparing the 2AFC procedure in Part I (Pre-test I) with a 2AFC procedure in which the order of channel presentations was randomized (Pre-testII). Note that in Pre-test I, the subject does not have to attend to the lights in order to identify the channel of information, while the Pre-test II procedure makes this attention necessary. Pre-test I was compared with Pre-test II in 5 test sessions, each containing 3 blocks of 100 Pre-test I trials and 3 blocks of 100 Pre-test II trials. This procedure provided a test of whether or not attending to the lights which identified the channels of information had any effect on the detection level of the subject.

Results. Table 9 contains the scores yielded by both subjects in the 5 test sessions of Part III. An analysis of variance was performed on the data from each subject separately, which determined the effects of Test Phase (I or II) and Test Session (1 - 5). Only the effect of Test Session for subject J was found to be significant ($F_{4,20} = 18.228$, $p < .01$). Tables 10 and 11 contain the results of both analyses of variance.

Part IV

Procedure. Given the apparent success of the threshold model in predicting the direction of results for both

Table 9

Pre-test I and Pre-test II Scores

for All Test Sessions of Part III

Test Session	B Test Phase	B Score	J Test Phase	J Score
I	I	74	I	70
	II	69	II	65
	I	73	I	54
	II	75	II	65
	I	71	I	60
	II	68	II	62
II	I	72	II	87
	II	75	I	77
	I	70	II	83
	II	65	I	79
	I	80	II	80
	II	69	I	85
III	I	78	I	81
	II	67	II	90
	I	61	I	84
	II	79	II	74
	I	62	I	78
	II	60	II	88
IV	II	72	II	80
	I	67	I	76
	II	68	II	78
	I	74	I	81
	II	71	II	83
	I	66	I	80
V	II	78	II	74
	I	71	I	78
	II	74	II	69
	I	76	I	73
	II	73	II	71
	I	80	I	67

N = 100

Table 10

Analysis of Variance of Subject B

Data in Part III

Source	SS	df	MS	F Ratio	F
A	188.2	4	47.05	MS_A / MS_E	1.5477
B	4.8	1	4.8	MS_B / MS_E	.1579
AB	36.8667	4	9.2167	MS_{AB} / MS_E	.3032
Error	608	20	30.4	———	———

A = Test Session
B = Test Phase

Table 11

Analysis of Variance of Subject J

Data in Part III

Source	SS	df	MS	F Ratio	F
A	1698.8667	4	424.7167	MS_A / MS_E	18.2282**
B	22.5444	1	22.5444	MS_B / MS_E	.9676
AB	20.4556	4	5.1139	MS_{AB} / MS_E	.2195
Error	466	20	23.3	———	———

A = Test Session
B = Test Phase

** = significant $p < .01$

subjects, additional analysis of the data was performed to test the model's ability to predict the magnitude of the results. From Green and Swets (1966) we see that the probability of detection, in the threshold model, based on n observations, is equal to one minus the product of the probability that detection will not occur on any of the n observations:

$$p_n = 1 - \prod_{i=1}^{n}(1 - p_i) \tag{1}$$

When it is assumed that all individual probabilities are equal, this formula becomes:

$$p_n = 1 - (1 - p)^n \tag{2}$$

Applied to the forced-choice task, with two response alternatives available (as is the case in this experiment), the probability of a correct response on a single trial becomes:

$$p(c) = 1 - (1 - p(D))^n + 1/2(1 - p(D))^n \tag{3}$$

where n is the number of observations of signal on a given trial and p(D) is the probability of detection on that trial. Table 12 contains the actual percentage of correct responses over all test sessions, along with the predicted probability of correct responses from the threshold model. P(D) was calculated from the pre-test and post-test data of each test session and this value was then used in equation (3) to obtain the predicted p(c) within each test session for n = 1, 2 or 3. This provided a test to see if the threshold model could predict from the subject's per-

Table 12

Predicted and Actual Probability
of Correct Responses
for all Test Sessions with χ^2 Comparisons

Subject/ Session	Est. p(D)	Pretest Posttest p(c)	n=1 Actual p(c)	n=1 Pred. p(c)	n=1 χ^2	n=2 Actual p(c)	n=2 Pred. p(c)	n=2 χ^2	n=3 Actual p(c)	n=3 Pred. p(c)	n=3 χ^2
B/1	.23	.615	.437	.615	39.48**				.69	.772	11.89**
B/2	.66	.83				.593	.942	1325.3**			
B/3	.54	.77	.69	.77	10.84**				.753	.951	244.28**
B/4	.59	.795	.52	.795	68.35**	.56	.916	499.59**	.687	.966	364.97**
B/5	.52	.76				.598	.885	484.47**			
B/6	.58	.79	.593	.79	36.59**	.67	.912	224.41**	.78	.963	126.56**
B/7	.28	.64	.547	.64	5.67*	.683	.741	5.0069*	.80	.814	.1757
J/1	.44	.72	.723	.72	.0165				.873	.912	6.064*
J/2	.13	.565				.633	.622	.3473			
J/3	.60	.80	.723	.80	11.02**				.807	.968	238.34**
J/4	.57	.785	.74	.785	1.946	.857	.908	8.86**	.853	.96	44.44**
J/5	.48	.74				.753	.865	64.07**			
J/6	.24	.62	.66	.62	1.019	.80	.711	11.80**	.88	.781	8.741**
J/7	.24	.62	.68	.62	2.29	.803	.711	12.69**	.90	.781	12.59**

* = significant $p < .05$
** = significant $p < .01$

formance on the pre-test and post-test (Part I) the results on the modified 4AFC task in Part II of the experiment.

<u>Results</u>. As revealed by the χ^2 comparisons between the actual and predicted probabilities of correct responses, shown in Table 12, the threshold model was not able to consistently predict the subject's level of performance. Accurate predictions would yield no significant differences between the actual and predicted p(c) data. However, 21 of the 30 χ^2 comparisons showed significant differences ($p < .05$). In almost all cases, the predicted values of p(c) were overestimates of the actual p(c) results.

Part V

<u>Procedure</u>. Despite the fact that the threshold model was not able to demonstrate reliable predictive ability in estimating the subjects' performance in Part II from the results in Part I of the experiment, further data analysis was conducted to determine whether or not the predicted relationship between probabilities of correct responses when n = 1, 2 or 3 still holds within a test session. That is, does equation (3) accurately predict the increase in p(c) as the number of observations of signal (n) within a trial increases from 1 to 3? Three test sessions (4, 6 and 7) incorporate trial sequences which yield all three

values of n within a test session. The data from these three test sessions for both subjects was converted into log scores to make a straight-line fit possible. Equation (3) was transformed into a straight-line equation in the following manner:

$$p(c) = 1 - (1 - p(D))^n + 1/2(1 - p(D))^n \tag{3}$$

$$= 1 - 1/2(1 - p(D))^n$$

$$2(1 - p(c)) = (1 - p(D))^n$$

$$-\log 2(1 - p(c)) = -n \log (1 - p(D)) \tag{4}$$

Equation (4), then, is a straight-line equation of the form $y = mx + b$ where:

$$y = -\log 2(1 - p(c))$$

$$x = n$$

$$m = -\log (1 - p(D))$$

$$b = 0$$

In theory, the straight line should go through the origin ($b = 0$). The least squares technique was used to fit the data from the three appropriate test sessions (4, 6 and 7) to this straight-line equation.

Results. Figure 2 (A - F) contains the best-fitting straight lines, along with the actual data points for each of the three test sessions for both subjects. It should be noted, that only one of the test sessions (i.e., session 4 for subject J) yielded results which deviated from the prediction that p(c) increases as n increases. All other

Figure 2 (A - F)

Least-squares Fit of Data from Sessions 4, 6 and 7 for Both Subjects to Straight-line Equation (4)

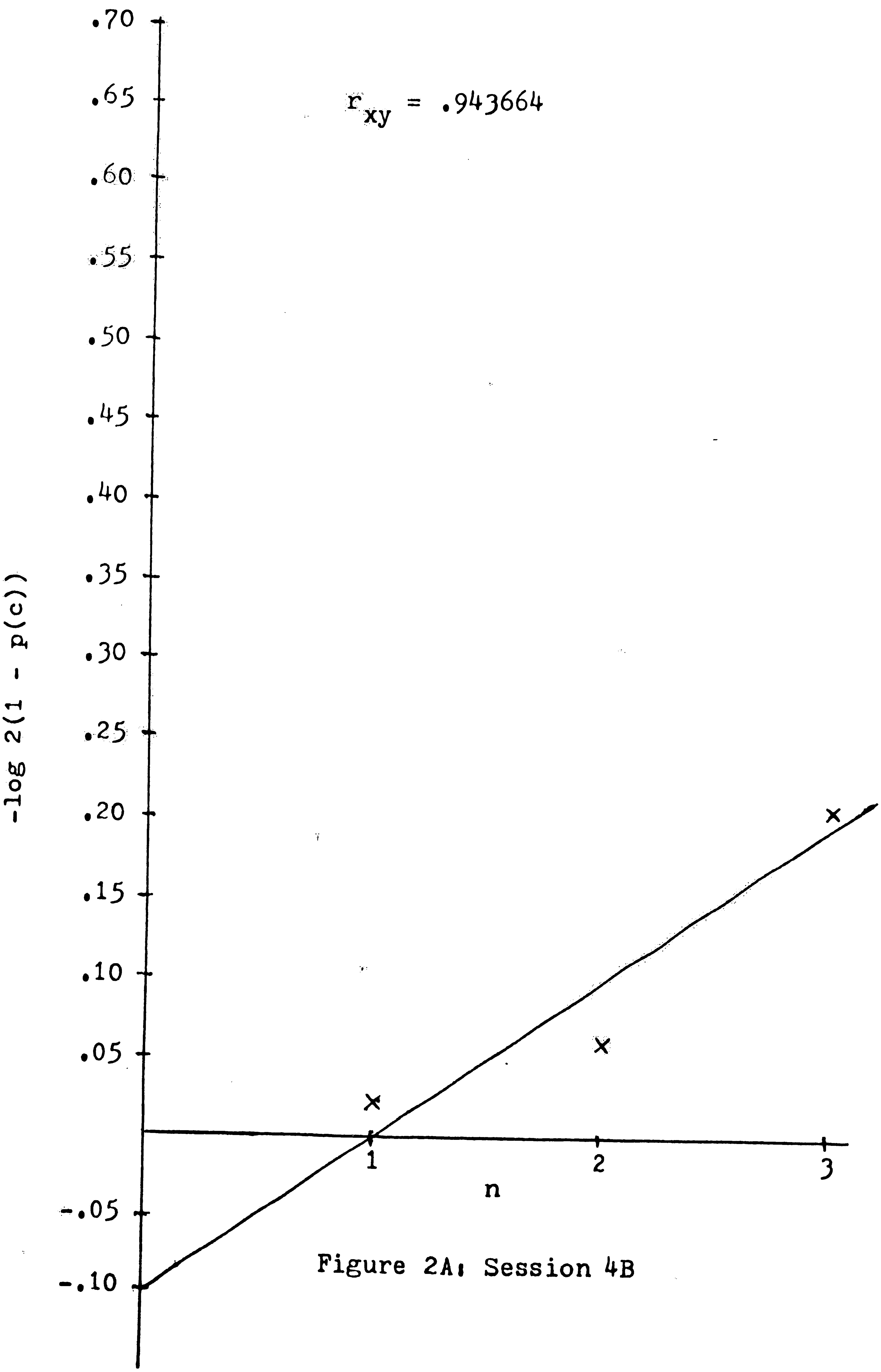


Figure 2A: Session 4B

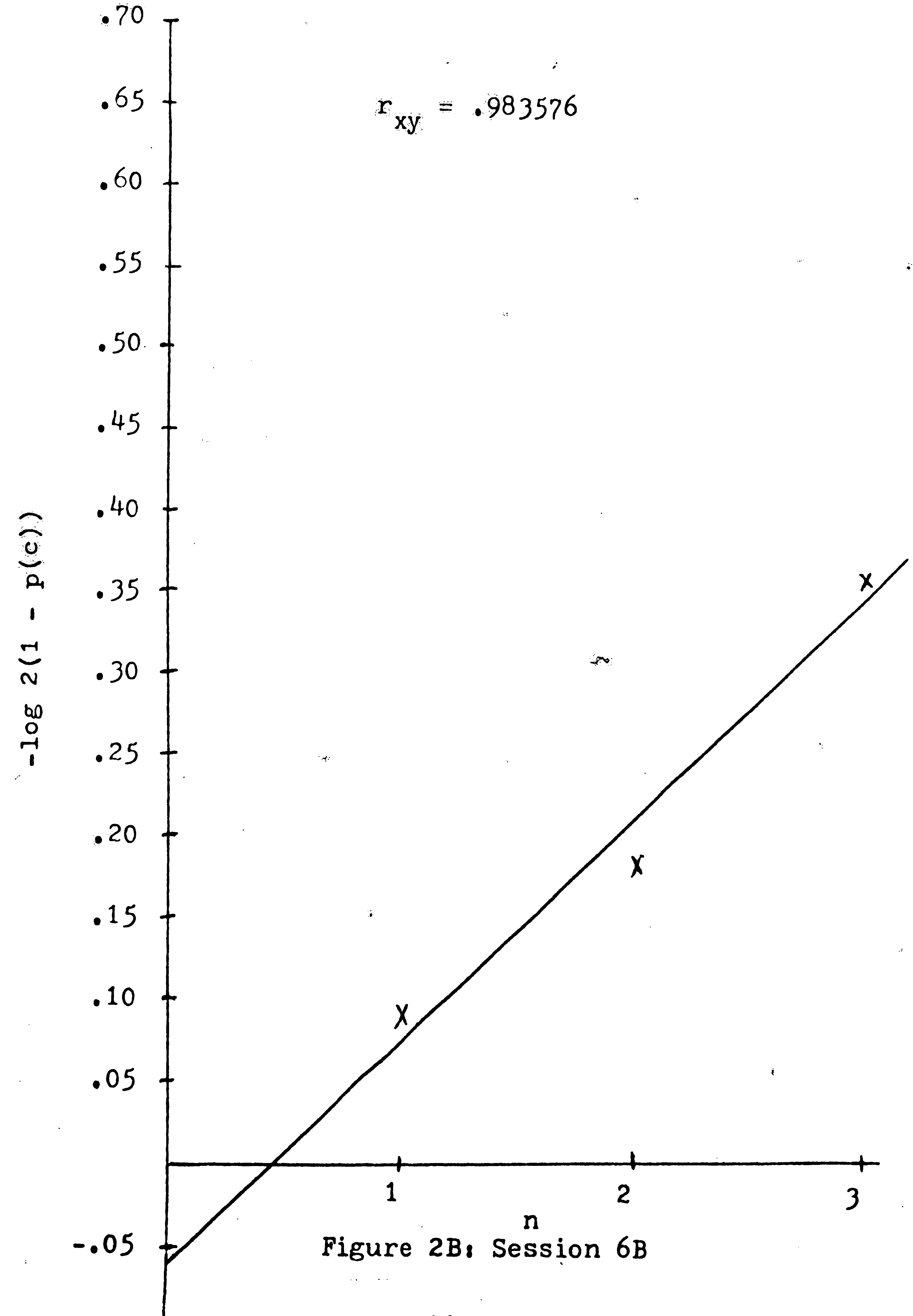


Figure 2B: Session 6B

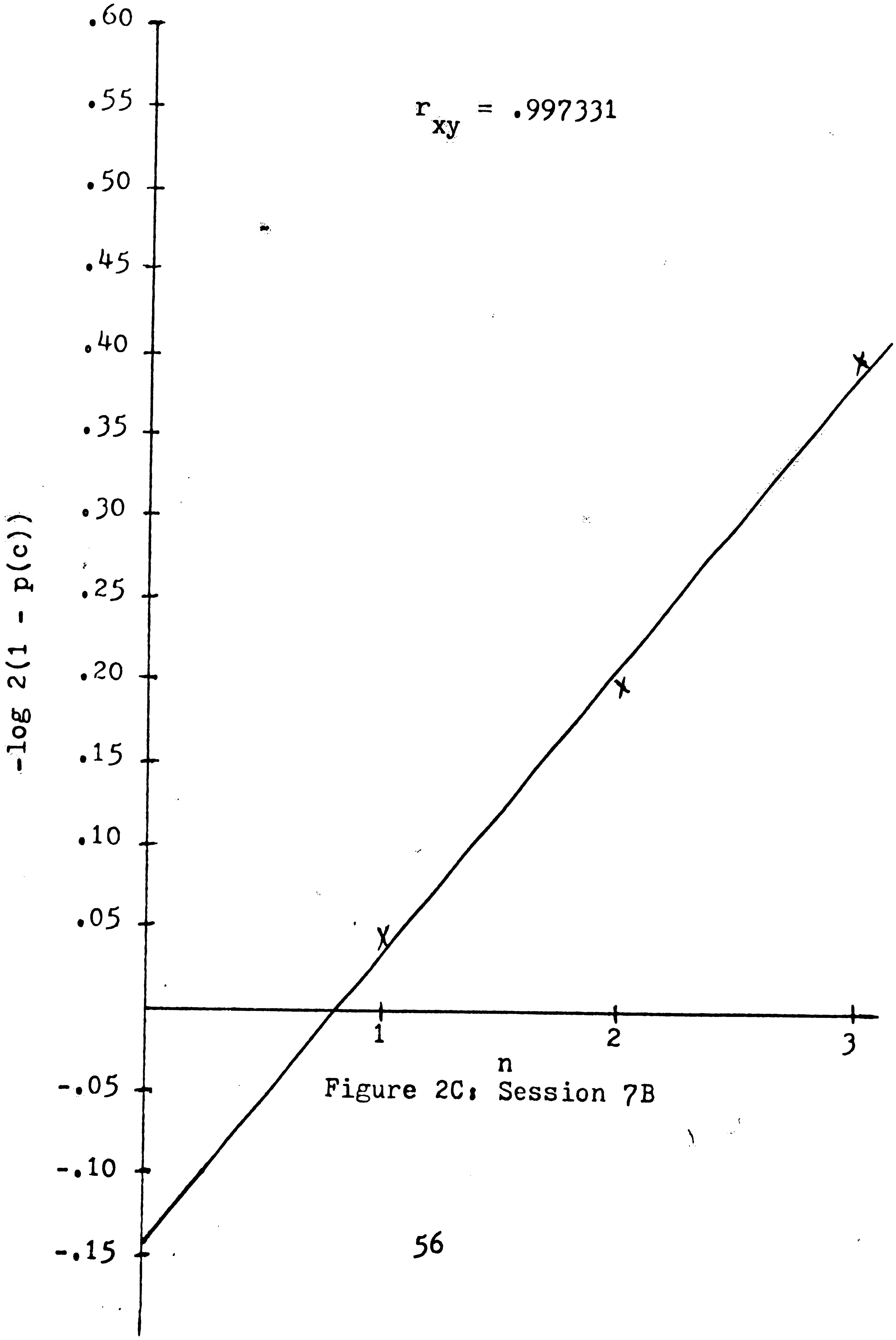


Figure 2C: Session 7B

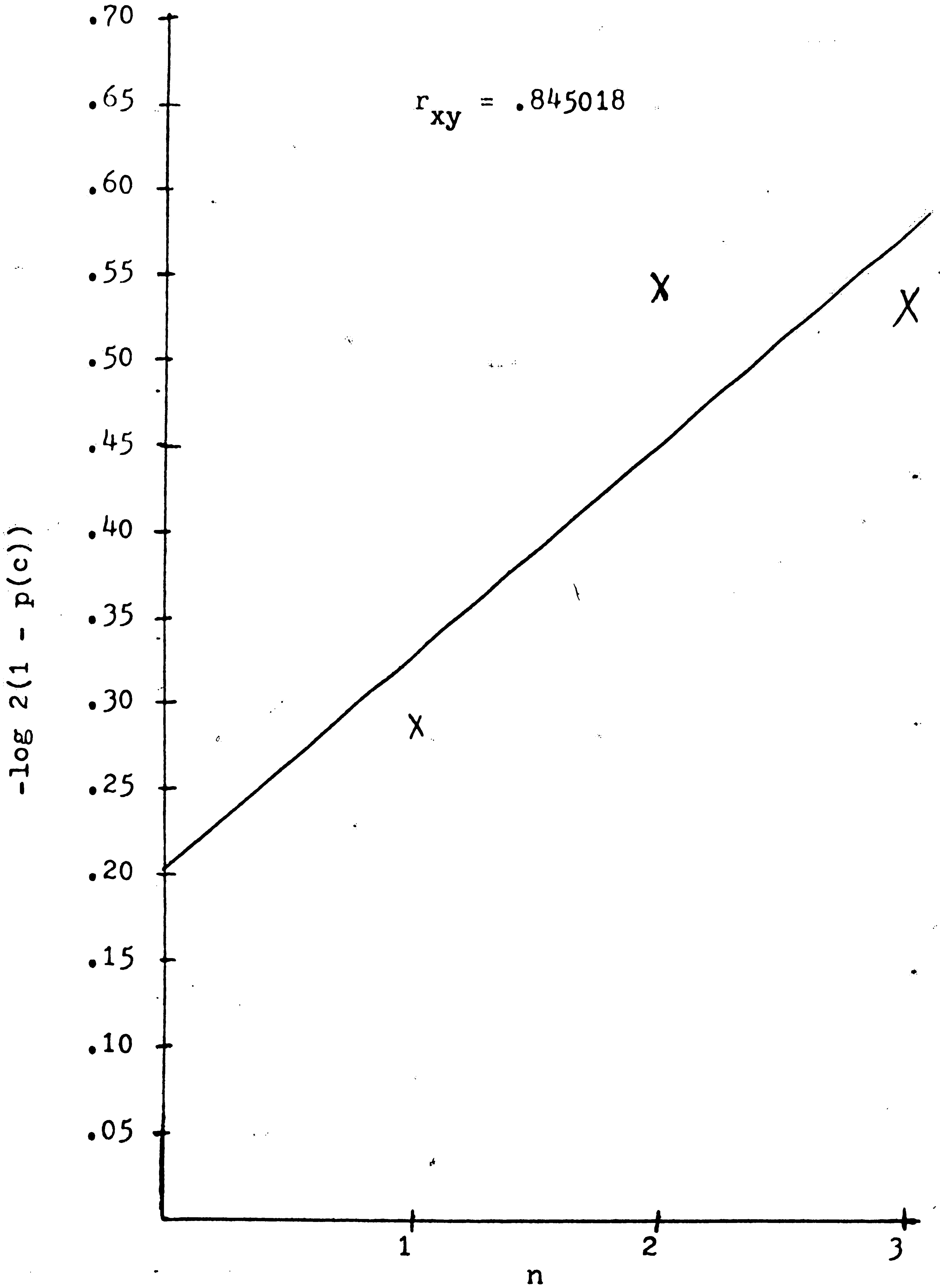


Figure 2D: Session 4J

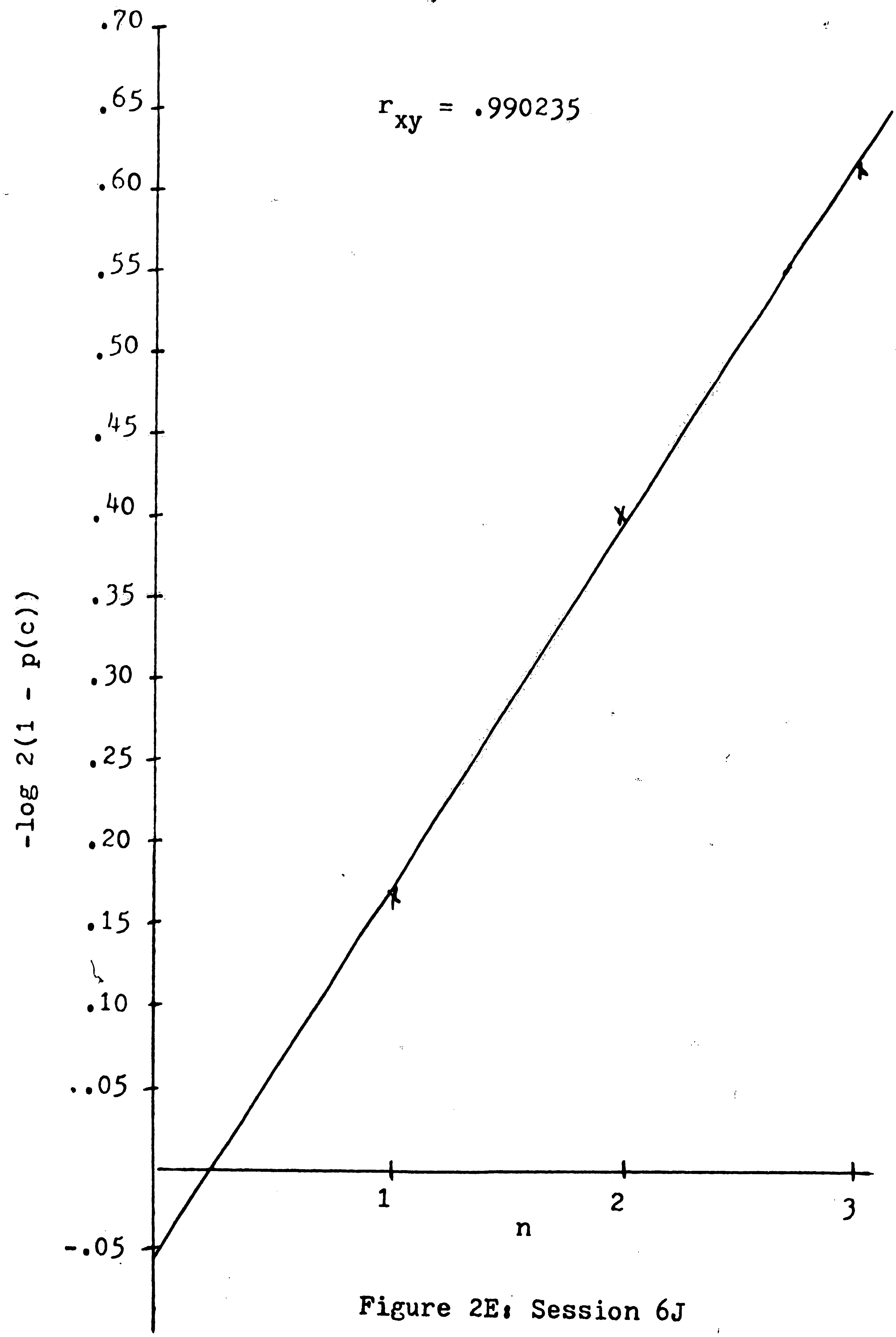


Figure 2E: Session 6J

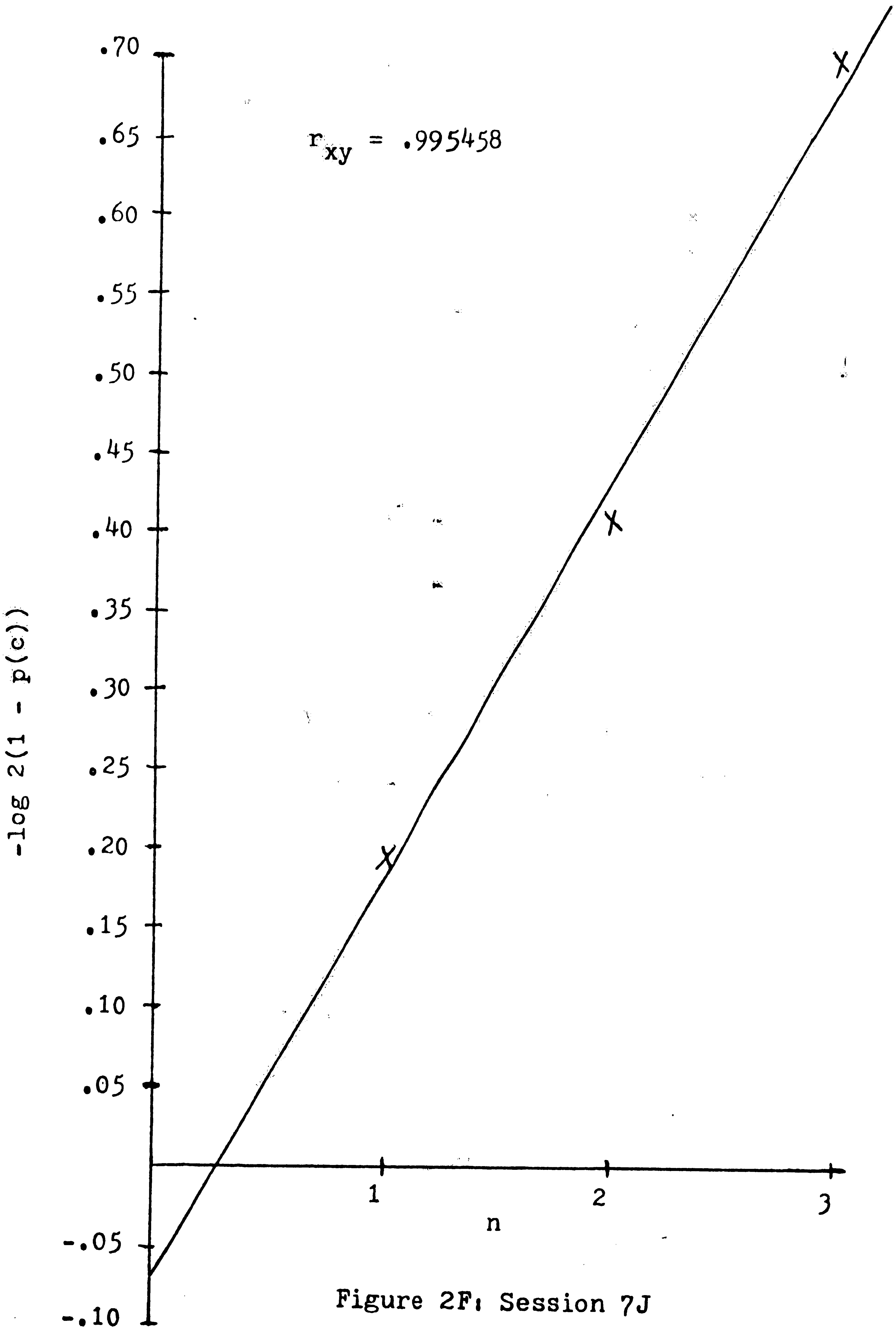


Figure 2F: Session 7J

sessions yielded results with high correlations between the actual data points and the straight line fit. Only one, however, from session B7, was significant ($p < .05$).

Discussion

The results obtained in Part I of the experiment are straightforward. With no significant differences between pre-test and post-test scores on any of the test sessions, it is reasonable to assume that the detection level for both subjects remained stable throughout the experiment. The use of trained observers, therefore, appears to have effectively eliminated threshold variability within a test session.

The results of Part II, although not conclusive, are very suggestive, and appear to lend strong support to the contention that both subjects were operating according to a threshold model in this detection task. Although the data presented in Table 7 suggest that some support was also given to the other models, this may be misleading. As indicated in the results of Table 8, on individual test sessions, some of the integration models make the same prediction as the threshold model. Therefore, although the overall results overwhelmingly support the threshold model, individual test session results will also support the other models as well.

In attempting to explain why the threshold model was supported in favor of the integration models, it may be helpful to re-examine some of the assumptions integration theory, as a whole, makes about the subject's receptor stage of detection. An important assumption of integration theory is that both S and N are integrable with equal ease. That is, the subject will receive just as much usable information from a presentation of N as he will from a presentation of S. In effect, this assumption implies that the subject can just as successfully detect a stimulus as "noise without signal" as he can "noise plus signal." This assumption, although mathematically convienent, may not be entirely accurate. Bell and Nixon (1971) tested this assumption using a YN task in which they asked their subjects to rate the stimulus presentations they received on the degree to which the waveform had signal or noise quality. That is, they responded on a 10 point scale how sure they were that the information they received was S (1 - 5) or N (1 - 5). The results obtained indicated greater reliability for ratings of S than for ratings of N. These investigators point to insufficient variations in the N waveform to permit meaningful rating of them. Whatever the explanation, there is evidence to support differences in observer ability to detect S and N presentations.

Another necessary and important assumption of integration theory is independence of observations. The information received from one observation, therefore, is not a factor of any previous observation. Again, however, this assumption may not hold due to a subject's tendency to establish some sort of response pattern. Speeth and Mathews (1961), using a 4AFC procedure, found dependencies between successive response intervals. That is, the subject's response depended, in part, upon his sequence of past responses. The length and degree of this dependency varied over subjects and with the particular sequence of intervals used, but it was shown to extend as far as 3 or 4 past intervals. These discrepancies between model assumptions and the actual behavior of subjects in a detection task may have strong consequences resulting in the lack of support shown for the integration models in this study. It should be noted, however, that the threshold model also assumes independence of observations. It is possible that any dependencies which resulted between observations in this task may have helped lead to good threshold model predictions.

Part III of the experiment attempts to show that the data from Part I are valid despite the fact that attending to the channel indicator lights was not necessary. In view of the results comparing the Pre-test I with the Pre-test II procedure, it can be stated that the two tasks do not yield differential results. The fact that subject J's results differ significantly over test sessions is unimportant, since all comparisons were made within a particular test session.

Part IV also yielded very conclusive results. Theoretically, the threshold model should be able to predict from the results of a 2AFC procedure, the subject's performance in a 4AFC procedure. However, the data in Part IV show that the prediction equation (3) for the most part yielded overestimates of the actual percentage of correct responses. There are a number of explanations for why this predicted relationship did not hold. Perhaps there is more than a quantitative difference between a 2AFC and a 4AFC detection task. The subject's response strategy could change when presented with these different tasks. Also, the term "modified" in the modified 4AFC procedure used in Part II may have had greater consequences on the results than expected. Instead of only one observation of signal in a 4AFC task, the subject could receive 1, 2 or 3 observations in the modified procedure with 2 rather than 4 response alternatives. Clearly, these procedures could differ in an unknown qualitative way, making theoretical comparisons cumbersome, if not impossible.

Part V data, however, show that relationships within the modified 4AFC procedure are consistent. That is, given a subject's performance when one observation of signal is observed ($n = 1$), his performance when $n = 2$ and $n = 3$ will follow from the relationship specified in equation (4). Although only one correlation is significant individually, the fact that 5 out of the 6 correlations exceed .94 is highly suggestive of the existence of a strong relationship. Since only 3 data points are used in forming the straight-line fit, the correlations are forced to be prohibitively high ($r = .997$, $p < .05$) in order to attain significance. Perhaps, due to the use of a modified rather than traditional 4AFC procedure, equation (4) does not specify the exact linear relationship in this particular case.

In conclusion, it is felt that the evidence presented strongly supports the threshold model's explanation for a subject's behavior in a signal detection task. More investigation is needed, however, before a definitive statement can be made regarding the predictive relationships which hold when using the tasks employed in this experiment. It would also be interesting to compare the same models using different types of tasks to see whether the success of the threshold model would generalize beyond the modified 4AFC procedure.

References

Bell, D. W. and Nixon, J. C. Reliability of Ratings in an Auditory Signal-Detection Experiment. Journal of the Acoustical Society of America, 1971, 49, 435 - 439.

Green, D. M. and Swets, J. A. Signal detection theory and psychophysics. New York: J. Wiley and Sons, 1966.

Gundy, R. F. Auditory Detection of an Unspecified Signal. Journal of the Acoustical Society of America, 1961, 33, 1008 - 1012.

Kay, E. J. The application of signal detectibility theory to a multiple-look task. Unpublished Ph.D. dissertation, Lehigh University, 1971.

Madigan, R. J. On the consistency of auditory judgements in signal detection. Dissertation Abstracts International, 1971, 31, 5031.

Maloney, D. M. and Welch, R. B. The effect of accessory auditory stimulation upon detection of visual signals. Psychonomic Science, 1972, 29, 345 - 347.

Robinson, D. E. and Trahiotis, C. Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions. Perception and Psychophysics, 1972, 12, 333 - 334.

Speeth, S. D. and Mathews, M. V. Sequential Effects in the Signal Detection Situation. Journal of the Acoustical Society of America, 1961, 33, 1046 - 1054.

Swets, J. A. Indices of Signal Detectability Obtained with Various Psychophysical Procedures. Journal of the Acoustical Society of America, 1959, 31, 511 - 513.

Swets, J. A., Shipley, E. A., McKey, M. J. and Green, D. M. Multiple observations of signal in noise. Journal of the Acoustical Society of America, 1959, 31, 514 - 521.

Vita

Joseph C. Weber, the son of Louise and Joseph Weber, was born in Bronx, New York on June 6, 1948. He attended All Hallows Institute from September, 1962 until June, 1966. He then entered Marquette University where he received a B.A. in Psychology in May, 1970.

In June, 1971 the author and Robin Mary Piper were married in North Lake, Wisconsin. After working for 1-1/2 years with emotionally disturbed adolescent boys at Milwaukee County Institutions, he entered Lehigh University in September, 1972. From then until the present, he has been a graduate student in the Psychology Department while holding a Teaching Assistantship.